Adaptation of arginine synthesis among uropathogenic branches of the Escherichia coli phylogeny reveals adjustment to the urinary tract habitat

Michael E Hibbing  
*Washington University School of Medicine in St. Louis*
Karen W Dodson  
*Washington University School of Medicine in St. Louis*
Vasilios Kalas  
*Washington University School of Medicine in St. Louis*
Swaine L Chen  
*National University of Singapore*
Scott J Hultgren  
*Washington University School of Medicine in St. Louis*

Follow this and additional works at: [https://digitalcommons.wustl.edu/open_access_pubs](https://digitalcommons.wustl.edu/open_access_pubs)

*Please let us know how this document benefits you.*

**Recommended Citation**

Hibbing, Michael E; Dodson, Karen W; Kalas, Vasilios; Chen, Swaine L; and Hultgren, Scott J, "Adaptation of arginine synthesis among uropathogenic branches of the Escherichia coli phylogeny reveals adjustment to the urinary tract habitat." mBio. 11, 5. e02318-20 (2020). [https://digitalcommons.wustl.edu/open_access_pubs/9752](https://digitalcommons.wustl.edu/open_access_pubs/9752)
Adaptation of Arginine Synthesis among Uropathogenic Branches of the *Escherichia coli* Phylogeny Reveals Adjustment to the Urinary Tract Habitat

Michael E. Hibbing, Karen W. Dodson, Vasilios Kalas, Swaine L. Chen, Scott J. Hultgren

*Department of Molecular Microbiology and Center for Women’s Infectious Disease Research, Washington University School of Medicine, St. Louis, Missouri, USA
Highest rank insignia
Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and Genome Institute of Singapore, Singapore

**ABSTRACT** Urinary tract infections (UTIs) are predominantly caused by uropathogenic *Escherichia coli* (UPEC). UPEC pathogenesis requires passage through a severe population bottleneck involving intracellular bacterial communities (IBCs) that are clonal expansions of a single invading UPEC bacterium in a urothelial superficial facet cell. IBCs occur only during acute pathogenesis. The bacteria in IBCs form the founder population that develops into persistent extracellular infections. Only a small fraction of UPEC organisms proceed through the IBC cycle, regardless of the inoculum size. This dramatic reduction in population size precludes the utility of genomic mutagenesis technologies for identifying genes important for persistence. To circumvent this bottleneck, we previously identified 29 positively selected genes (PSGs) within UPEC and hypothesized that they contribute to virulence. Here, we show that 8 of these 29 PSGs are required for fitness during persistent bacteriuria. Conversely, 7/8 of these PSG mutants showed essentially no phenotype in acute UTI. Deletion of the PSG *argI* leads to arginine auxotrophy. Relative to the other arg genes, *argI* in the B2 clade (which comprises most UPEC strains) of *E. coli* has diverged from *argI* in other *E. coli* clades. Replacement of *argI* in a UPEC strain with a non-UPEC *argI* allele complemented the arginine auxotrophy but not the persistent bacteriuria defect, showing that the UPEC *argI* allele contributes to persistent infection. These results highlight the complex roles of metabolic pathways during infection and demonstrate that evolutionary approaches can identify infection-specific gene functions downstream of population bottlenecks, shedding light on virulence and the genetic evolution of pathogenesis.

**IMPORTANCE** Uropathogenic *Escherichia coli* (UPEC) is the most common cause of human urinary tract infection (UTI). Population bottlenecks during early stages of UTI make high-throughput screens impractical for understanding clinically important later stages of UTI, such as persistence and recurrence. As UPEC is hypothesized to be adapted to these later pathogenic stages, we previously identified 29 genes evolving under positive selection in UPEC. Here, we found that 8 of these genes, including *argI* (which is involved in arginine biosynthesis), are important for persistence in a mouse model of UTI. Deletion of *argI* and other arginine synthesis genes resulted in (i) arginine auxotrophy and (ii) defects in persistent UTI. Replacement of a B2 clade *argI* with a non-B2 clade *argI* complemented arginine auxotrophy, but the resulting strain remained attenuated in its ability to cause persistent bacteriuria. Thus, *argI* may have a second function during UTI that is not related to simple arginine synthesis. This study demonstrates how variation in metabolic genes can impact virulence and provides insight into the mechanisms and evolution of bacterial virulence.

**KEYWORDS** *Escherichia coli*, arginine metabolism, positive selection, urinary tract infection
Urinary tract infections (UTIs) are among the most common bacterial infections of humans, affecting over half of all women and accounting for ~$2.5 billion in annual health care-related expenses in the United States (1, 2). Uropathogenic *Escherichia coli* (UPEC) accounts for the majority of community-acquired UTIs, typically acute uncomplicated cystitis (2, 3). Unfortunately, after an initial acute UTI, ~25% of women experience at least one episode of recurrent UTI (rUTI) within 6 months of the initial infection (4, 5). While in some patients acute UTIs resolve within a week, other patients can develop chronic recurrent cystitis even despite treatment (6). Further, frequent empirical antimicrobial treatment of acute UTIs and prophylactic antimicrobial treatment to prevent rUTIs has resulted in the emergence and global dissemination of multidrug-resistant UPEC (7–9).

Experimental mouse models of UTI are able to mimic many of the known clinical manifestations of human UTIs, including acute cystitis, recurrent UTI, and chronic cystitis (10–14). During acute UTI, UPEC employs the FimH adhesin of the type 1 pilus to invade urothelial superficial facet cells via an endocytic mechanism (10–15). Following invasion, some UPEC organisms evade host cell exocytic mechanisms, escape the endocytic vesicle, and rapidly replicate to form dense clusters (~1,000 to 10,000 CFU/epithelial cell) of clonal cytoplasmic bacteria, termed intracellular bacterial communities (IBCs) (16–21). IBCs are resistant to antibiotic treatment and the host immune response. In addition, quiescent intracellular reservoirs (QIRs) can form in the underlying transitional epithelial cells, comprising clusters of 2 to 12 membrane-bound bacteria that are dormant and thereby tolerant of antibiotics and host defenses (22–24). QIRs can persist for weeks to months, during which time the urine may be sterile; however, bacteria can reemerge from a QIR to cause a rUTI (25–27). Importantly, structures similar to IBCs can be found in the urine of human patients suffering from UTI, further supporting the relevance of experiments with the mouse model (28–31).

Mice with a history of UTI are predisposed to chronic rUTI. In mice, chronic cystitis occurs downstream of the IBC bottleneck and is defined as an extracellular infection characterized by persistent, high-titer bacteriuria (>10⁴ CFU/ml) and bladder colonization (>10⁴ CFU/bladder), chronic pyuria, ablation of the terminally differentiated superficial cells, urothelial hyperplasia, lymphoid follicles, and urothelial necrosis persisting for weeks (10). Mice with a history of chronic cystitis can be cured by antibiotic treatment; however, they are thereafter sensitized to higher rates of rUTI in subsequent challenge infections due to bladder remodeling (10, 13, 14). In mice, infection can leave a molecular imprint resulting in hundreds of differentially expressed genes compared to the bladders of naive mice; this physiological remodeling results in a predisposition to rUTI (13). Of note, in humans, the history of UTI is one of the most significant risk factors predisposing to a rUTI (32, 33).

One of the main experimental hurdles in studying factors important in persistent UTI and rUTI is the progressive population bottleneck that begins during the initial acute phase of cystitis. The bacteria within an IBC are clonal (each IBC is descended from a single invasive UPEC organism), and only 1 of 10⁵ to 10⁶ experimentally inoculated UPEC organisms forms an IBC (16). Therefore, IBC formation represents one severe population bottleneck that occurs during acute infection (16). The formation of QIRs and development of chronic cystitis incur additional bottlenecks, as measured by random loss of coincoculated, isogenic, differentially marked strains (16). The mechanisms causing these bottlenecks are not fully defined, but they are likely related to some combination of bacterial invasion, immune cell killing, intracellular innate immune responses, competitive growth, and continuous micritution. Because modern unbiased genetic approaches (like signature-tagged mutagenesis or transposon insertion sequencing [Tn-seq] screens) require testing of high-complexity mutant libraries (i.e., those having a large number of genetically distinct clones), population bottlenecks (which cause random loss of clones) prevent identification of mutants that exhibit defects downstream of the bottleneck (34–36). Recently, unbiased global approaches to interrogate UPEC, including comparative genomics, transcriptomics, and proteomics, have also been applied to
analyze factors required for UTI pathogenesis (37–41), though these have not addressed chronic rUTI.

In silico sequence-based methods provide an alternative approach to identifying virulence factors that does not suffer from experimental population bottlenecks. UPEC are hypothesized to be better adapted to the urinary tract than non-UPEC strains. Any adaptive mutation(s) should be evolving under positive selection; therefore, we and others have conceptually connected loci under positive selection with virulence-related genes (42–49). Indeed, FimH, a critical virulence factor for acute UTI and IBC formation, is evolving under positive selection in UPEC, and identification of individually selected amino acids has led to fundamental insights into the structure, function, and evolution of FimH and the importance of a finely tuned conformational equilibrium to its function in UTIs (48–50). Thus, positive-selection analysis can identify virulence factors (and their functions) that otherwise might be difficult to study due to population bottlenecks.

We previously used an in silico comparative genomics analysis to identify 29 genes evolving under positive selection (here referred to as positively selected genes [PSGs]) specifically in UPEC but not other types of E. coli (47). These 29 PSGs are predicted to contribute to a diverse set of cellular functions, including iron uptake, membrane protein localization, DNA structure and repair, small-molecule import/export, cell division, and arginine metabolism (47); however, their roles in chronic UTI have not been characterized. Here, we tested the impact of these 29 PSGs on the fitness of UPEC in a mouse model of UTI, focusing on chronic cystitis (10–12, 51). In total, single gene deletions of 10 different PSGs had lower fitness in at least one of our in vivo infection models, with 8 single PSG deletions showing defects during chronic cystitis. Further analysis of the argI gene (encoding the anabolic ornithine transcarbamoylase) showed that while arginine biosynthesis is important for in vivo fitness, argI also seems to have an additional in vivo function during UTI. Specifically, an allelic replacement of a clade B2 argI with a non-B2 argI allele complements arginine auxotrophy but not in vivo virulence. In keeping with this, argI from the B2 clade, which comprises the majority of UPEC strains, is highly diverged from the argI genes found in other E. coli clades (39,52, 53). Together, these results reveal an unexpected complexity in the importance of arginine during chronic UTI, which could not have been found from traditional genomic mutagenesis studies due to the population bottlenecks occurring during in vivo UTI. They also highlight the value of positive-selection analysis, which has now identified in vivo-specific roles mediated by allelic variation in both fimH and argI. Finally, our results suggest that screening for anomalous divergence patterns, a rapid technique amenable to modern genomic data sets, may be a useful approach to identify other genes involved in virulence and bacterial evolution.

RESULTS

**PSG mutations have little effect on in vitro phenotypes.** The primary hypothesis is that genes under positive selection in UPEC play a role during UTI in vivo. We created knockouts in each of the 29 previously identified PSGs in UTI89, a prototypical UPEC clinical isolate. We first tested these mutants for in vitro phenotypes previously shown to be associated with pathogenesis: (i) type 1 pilus expression and function as measured by guinea pig red blood cell hemagglutination (HA) assay and (ii) growth in rich (LB) medium and in pooled filtered human urine (54–56). All the PSG mutants were indistinguishable from the wild type in all assays with two exceptions (Fig. 1A to C): (i) the ΔrecC mutant had a general growth defect in both LB and urine (Fig. 1B and C), and (ii) the ΔargI mutant had a mild growth defect in urine (but not in LB), wherein the cell density during stationary phase was lower (Fig. 1B and C). Supporting these observations, deletion of either recC or argI in the UPEC isolate CFT073 was recently shown to cause an in vitro growth defect in human urine (36).

**Multiple PSGs have a role in chronic UTI.** We next tested the competitive fitness of each PSG mutant in a mouse model of chronic UTI, in which naive female 7- to 8-week-old C3H/HeN mice were inoculated with a mixture of ~10⁸ CFU each of a PSG mutant and wild-type UTI89. Infected mice were monitored by urinalysis until 28 days...
post-infection (dpi), at which time the mice were sacrificed and bladder and kidney burdens were determined (Fig. 2A and Fig. S1 and S2). It was previously shown that these infections result in two different UTI outcomes: (i) high-titer and chronic infection and (ii) resolution with or without the formation of QIRs (10). These outcomes are reliably distinguished by persistently high urine titers (>10^4 CFU/ml), indicating chronic infection, and low or variable urine titers, indicating resolution; we therefore analyzed the data for mice with chronic and resolved infections separately (Fig. 2A and Fig. S1 and S2). Eight PSG mutants had chronic cystitis defects (ompC, nepI [previously yicM (57)], recC, entF, cutE, yjbN, mdtC [previously yegO (58)], and argI), with all being dramatically outcompeted by the wild type, with median log_{10}(CI) values below -4 at 28 dpi (indicating that there are >10,000 times more wild-type than mutant bacteria) (Fig. 2A). Six PSG mutants had fitness defects in the resolved group (amiA, entD, cutE, yjbN, mdtC, and argI), though due to the lower overall titers, the median log_{10}(CI) values ranged between -0.2 and -1.29 at 28 dpi (Fig. S1 and S2). Notably, as strains with mutations in 4 PSGs (cutE, yjbN, mdtC, and argI) had defects in mice with chronic and resolved infections, there were a total of 10 PSGs that contributed to some form of long-term persistence in the urinary tract (Fig. 2A and Fig. S2). Interestingly, single-infection assays showed that, when inoculated individually into mice, all PSG mutants tested achieved bacterial titers indistinguishable from that of the wild type at 28 dpi, again with the exception of the ΔrecC mutant (Fig. 2B).

**PSG mutants affecting chronic infection have little effect on acute UTI.** The longitudinal urinalysis conducted during the competitive chronic infections revealed that the median log_{10}(CI) of the 8 PSG mutants with chronic cystitis defects were generally close to 0 for the first 24 h (i.e., indistinguishable from that of the wild type) (Fig. S1). We therefore tested the fitness of the 8 mutant strains in acute-cystitis experiments: single infections at 6 hpi and competitive infections at 18 hpi (Fig. 2C and D). Six of the eight mutants were indistinguishable from the wild type in these acute
infections. The most pronounced exception was the ΔrecC mutant, which showed a substantial defect in both assays (and which was also the mutant with the most severe in vitro growth defects). Furthermore, consistent with having a mild growth defect in vitro in urine, the argI mutant had a statistically significant, but quantitatively small, defect in these acute-infection assays. Interestingly, it was recently found that, in CFT073, deletion of either recC or yojI resulted in an acute fitness defect (36).
**De novo arginine synthesis is required during competitive chronic UTI.** Several studies have suggested a role for arginine metabolism during UPEC UTI (37, 59–63). However, arg1 is the only arg gene under positive selection in UPEC. To further investigate the role of arginine metabolism and the function of Arg1 during UTI, we generated individual strains with mutations in argA and argG, two genes predicted to encode metabolic enzymes that come biochemically before and after Arg1, respectively, in the arginine biosynthetic pathway (Fig. 3A and Fig. S6) (47, 64). These genes are each in monocistronic operons, and thus, their mutation is not predicted to induce polar effects. An ΔargA mutant should have arginine synthesis disrupted upstream of the substrate for Arg1, while an Δargl mutant should have synthesis disrupted upstream of the substrate for ArgG (Fig. 3A and Fig. S6) (47, 64). In agreement with previous work in CFT073, all three arg mutants had similar phenotypes in vitro, exhibiting arginine auxotrophy in minimal defined media and a lower cell density at entry into stationary phase when grown in filtered pooled human urine (Fig. 3B and Fig. S6) (36). In particular, high-cell-density growth in urine was restored by supplementation of arginine or of the biosynthetic intermediates downstream of the mutated enzymatic step (Fig. 3C and Fig. S6). We then tested the argA and argG mutants in competitive chronic cystitis infections. Both the ΔargA and ΔargG mutants also had significant defects. In particular, the ΔargG mutant was similar to the Δargl mutant, in that they were rapidly outcompeted by wild-type UTI89 in the urine samples [median log10(CI) > 4.0 by 14 dpi] and had median log10(CI) values from the bladders that were below −4.0 in mice at 28 dpi (Fig. 3D to F).
B2 strains of *E. coli* have a divergent allele of *argI*. The above-described experiments with the Δ*argA* and Δ*argG* mutants argue that the role of ArgI in arginine biosynthesis does not explain why it is evolving under positive selection. To understand whether allelic variation (as opposed to deletion mutants) might provide further insights (as it did with *fimH*), we assessed the conservation of the *arg* genes via a pairwise comparison of arginine anabolic protein sequences from UTI89 and the commensal *E. coli* isolate MC4100 (Fig. S8). In this comparison, the ArgI sequences had 20 variant residues out of 334 total amino acids (94% identity), while the other arginine anabolic proteins were 98 to 100% identical between these isolates (Fig. S8). Interestingly, the majority of the variable amino acids were spatially far from both the enzymatic active site and the trimer interfaces (Fig. S8). We then examined the same *arg* genes in a set of 2,274 publicly available *E. coli* genomes (all NCBI RefSeq assemblies as of 26 April 2016). Again, *argI* was unique in that there was a largely B2-specific clade of *argI* alleles which were separated from the other *E. coli* *argI* alleles by a relatively long evolutionary branch (Fig. 4A). While B2-specific alleles were found in other *arg* genes,
they were not as clearly divergent from the other *E. coli* alleles. (Fig. 4B). The conserved argI allele in the B2 clade of UPEC was significantly divergent (lowest nucleotide and predicted amino acid identity) from argI alleles in the other *E. coli* clades. This demonstrates that there is a conserved argI allele in the B2 clade, which was divergent from argI in the other B2 clades. (Note that while sequence differences are required for positive-selection detection algorithms to work, the existence of positive selection does not necessarily correlate with sequence diversity in a given gene.)

A B2 allele of argI has an additional *in vivo* function during competitive chronic UTI. To assess whether argI has an additional role during UTI aside from arginine biosynthesis, we generated allelic replacement strains of UTI89, reintegrating the UTI89 (a B2 strain) or MC4100 (a non-B2 strain) allele of argI (designated argI\textsubscript{UTI89} and argI\textsubscript{MC4100}, respectively) into the original argI locus in UTI89 using a scarless cloning technique (65). As expected, both argI alleles complemented all *in vitro* growth phenotypes (arginine auxotrophy in minimal media and low-density transition to stationary phase in urine) (Fig. S8). To further test for subtle fitness defects, we performed competitive *in vitro* growth assays between strains carrying either allele; again, no differences were detectable between otherwise isogenic strains carrying these two argI alleles (Fig. S8).

We then performed competitive, *in vivo* chronic infections between these strains carrying the argI\textsubscript{UTI89} and argI\textsubscript{MC4100} alleles. These infections were performed as described above, except they were allowed to proceed longer (63 dpi), as an initial pilot experiment suggested that the full effect of the non-B2 argI allele had not manifested by 28 dpi (data not shown). Remarkably, despite no detectable differences *in vitro*, bacteria carrying argI\textsubscript{UTI89} outcompeted those with the argI\textsubscript{MC4100} allele in the urine, bladders, and kidneys of chronically infected mice, reaching a median log\textsubscript{10}(CFU) of −5.4 in the bladder at sacrifice at 63 dpi (Fig. 5 and Fig. S9). The strain carrying argI\textsubscript{MC4100} was nearly undetectable at this time. Importantly, at 28 dpi (matching the time point for previous infections), the argI\textsubscript{MC4100} strain was already significantly outcompeted by the argI\textsubscript{UTI89} strain in urine (Fig. 5).
DISCUSSION

Different *E. coli* strains are known to differ in their propensity to cause disease in humans (and other organisms), giving rise to disease syndrome-based terms such as UPEC, enterohemorrhagic *E. coli* (EHEC), enteropathogenic *E. coli* (EPEC), avian pathogenic *E. coli* (APEC), and neonatal meningitis-causing *E. coli* (NMEC), among others. In other words, some *E. coli* strains are better adapted to survive in different habitats. A firm genetic definition of these various types of pathogenic *E. coli* remains elusive; however, most of the specific disease phenotypes have been attributed to known or putative virulence factors, although many of these have not been fully experimentally verified. UPEC, in particular, is a heterogeneous group of *E. coli* strains; while most are found within the B2 and D clades, there remains substantial variation in gene content even among B2 clade UPEC strains and no clear genetic signature of UPEC (66, 67). One approach to identifying genetic adaptations that enable strains to cause urinary tract infections (or other diseases) is to look for loci under positive selection (47, 68).

Positive selection is an evolutionary process whereby organisms with genetic changes that increase fitness are more likely to survive or reproduce, thus increasing the frequency of those changes in the population with time. In the context of virulence, the genetic factors that enable some strains to cause a specific disease are those that increase fitness and therefore should be evolving under positive selection. In the specific case of UPEC, a direct connection between positive selection and virulence-related genes has been posited (47, 68). One of the best-characterized examples in UPEC is the *fimH* gene, which encodes the FimH tip adhesin that tips the type 1 pilus. FimH has a critical role in the establishment of UTI as the primary binding determinant for UPEC to specifically attach to and invade bladder epithelial cells by engaging mannosylated surface proteins (17, 20, 69). The *fimH* gene is evolving under positive selection (43, 46, 48, 68). Interestingly, the residues detected as increasing fitness initially had no detectable *in vitro* role in binding to mannose (48). Subsequent work showed that the amino acids under positive selection in FimH affected the dynamics of switching between high- and low-affinity mannose binding states, which was a previously unappreciated step during UTI (49, 50).

In this study, we comprehensively tested a set of 29 additional PSGs that were previously identified by comparing sequences of UPEC versus non-UPEC *E. coli* strains. Deletion of 8 of these genes lowered fitness in a mouse model of chronic UTI. This model mimics persistent infections experienced by some human patients. In this model, high titers of UPEC are continuously present in the urine for the lifetime of the animal or until the infection is cleared by antibiotics, implying that there may be a strong selection for utilization of available nutrients (10, 12, 13). Interestingly, PSG mutants demonstrated infection defects only when competing with an otherwise isogenic wild-type strain. In humans, UTIs likely arise from the introduction of an inoculum consisting of mixed bacterial species and *E. coli* strains from gastrointestinal and vaginal reservoirs into the urinary tract (70–73). At the time of diagnosis, however, UTIs are generally clonal infections consisting of high bacterial loads of a single strain (74–76), implying that naturally occurring human UTIs may, in fact, also be competitive infections. This competition will occur between *E. coli* and other bacteria in the inoculum, which should include other members of the vaginal or perineal flora (such as *Lactobacillus*); however, multiple strains of individual species (particularly *E. coli*) could also be present in the inoculum, either closely or distantly related to one another, mimicking the competitive experimental infections we have studied here.

The importance of growth and arginine metabolism led us to perform further studies on *argI* (37, 59–63). Mutation of *argI* abolishes the ability of UPEC to synthesize arginine, leading to auxotrophy for arginine. To test the hypothesis that arginine metabolism is important in chronic UTI, we made deletion mutations in two other genes, *argA* and *argG*, which also result in arginine auxotrophy. As expected, these mutations also led to a fitness defect in chronic UTI. These results demonstrate that amino acids, in particular arginine, are present at low levels in urine and that the ability
to synthesize them de novo confers a competitive advantage to UPEC during UTI, but they do not explain why argI is the only arginine synthesis gene evolving under positive selection.

We therefore reexamined the sequence of argI and other arginine metabolic genes in a broader set of ≈2,200 E. coli genome sequences, which indicated, consistent with previous reports, that argI is evolving under positive selection and has specifically diverged in B2 strains of E. coli. This divergence has a specific functional consequence in vivo, as a UPEC strain carrying either a B2 allele or an A allele (a clade of E. coli containing many commensal GI strains) has no auxotrophic phenotype for arginine; however, the strain carrying the B2 allele outcompetes an isogenic strain carrying the A allele as if the latter were an argI knockout. Interestingly, the clade D argI allele of E. coli was not under positive selection, despite the fact that E. coli strains from this clade are frequently observed causative agents of UTI (39). This finding suggests that there may be multiple solutions that E. coli has evolved to facilitate persistence in the urinary tract. We suggest that further positive selection analysis focusing on clade D E. coli might reveal aspects of such an alternative mechanism.

These results with argI form a strong parallel with those previously observed for fimH. Both genes have a clear function that is readily measured in vitro and known to be important for UTI in vivo. In addition, both genes have allelic variants naturally found among E. coli that do not obviously affect their function in vitro but result in strong fitness defects in vivo. Similarly, it was recently shown that the oxidative fumarase FumC is essential for fitness during UTI pathogenesis, but this essentiality appears to be predominantly associated with the role FumC plays in coping with iron limitation, via its ability to function enzymatically without an Fe-S cluster, rather than its specific role in the tricarboxylic acid (TCA) cycle (77). We therefore hypothesize that, like with FimH and FumC, a potential second, and possibly more subtle, function not currently appreciated in vitro exists for ArgI. This second ArgI function would have specifically evolved in B2, and its importance during chronic UTI would account for its being under positive selection (though other environments or disease states may select for it as well). The arginine synthesis step catalyzed by argI is situated at a branch point where biosynthetic precursors can be shunted from arginine synthesis into the biosynthetic pathway for several polyamines. The flux through the synthesis pathways of arginine and of polyamines, which are multifunctional compounds that have been shown to contribute to stress responses in UPEC, represent one potential mechanism by which the B2 allele of argI contributes to fitness (61, 78, 79). Conversely, the enzymatic steps catalyzed by ArgA and ArgG are relatively distal from this branch point and would likely have less impact on the relative levels of arginine, ornithine, and the other arginine-derived polyamines, suggesting one possible reason why argI is under positive selection in clade B2 E. coli but argA and argG are not.

We therefore conclude that positive selection analysis is indeed a powerful and complementary technique for identifying genes important for virulence. The success rate of PSG validation (deletion of 10/29 PSGs has a fitness defect, during chronic infection [8 PSGs] and/or QIR formation [6 PSGs]) compares favorably to other methods for identifying UPEC virulence genes, such as gene presence/absence analysis, candidate gene analysis (mostly targeted at “traditional” virulence factors or pathogenicity islands), and transcriptomics analyses (38, 40, 80–84). Of note, other mutant screens, such as Tn-seq, have very high validation rates. This is likely due to the fact that the “putative virulence factors” identified in the initial large-scale mutagenesis Tn-seq library screens are most often validated using the same pathogenicity model under which they were first selected. PSG analysis is more similar to a bioinformatic genomic comparison to determine presence/absence of genes in particular pathogenic strains compared to nonpathogenic strains. These putative virulence genes and/or specific gene alleles determined bioinformatically may be selected for in any of a number of different environments or conditions under which these strains reside and/or cause infection. Thus, the success of validating the role of any genomically identified candidate virulence factors is dependent upon which model they are tested in. We tested our
PSGs in a UTI model and found that strains with deletions in any of 10 of them had measurable UTI phenotypes. It may be possible that the other 19 PSGs for which we found no UTI phenotype may provide a selective advantage in other habitats related to UTI conditions, residence in the gastrointestinal tract (GIT), or vaginal reservoirs (85, 86). Further, the particular alleles of PSGs may have phenotypes that cannot be elucidated by testing deletion mutants but may require testing using allele swaps.

In addition, as shown for fimH and now argI, positive selection analysis potentially captures alternative gene functions that are specifically required in vivo. More generally, positive selection analysis is applicable to any infectious agent, sequences of which are available at exponentially increasing rates. In particular, positive selection may be especially useful for studying pathogenic processes that, like UTI, have stringent bottlenecks and strong founder effects, such as those of HIV, *Salmonella enterica* serovar Typhimurium, and *Borrelia burgdorferi* (87–90).

**MATERIALS AND METHODS**

**Ethics statement.** All animal experiments were conducted in accordance with the National Institutes of Health guidelines for the housing and care of laboratory animals as well as institutional regulations after pertinent review and approval by the Animal Studies Committee at the Washington University School of Medicine (protocol number 21080276). Human urine used for growth determination was collected from healthy consenting volunteers according to Hultgren lab human studies protocol number 201207143.

**Media, reagents, and mutant generation.** The UPEC strains used in this study are all derivatives of the human cystitis isolate UTI89 and are listed in Table S1 (17, 91). In general, bacteria were grown and propagated in Luria-Bertani (LB) broth and plated for isolation on LB agar (BD). Where necessary, the medium was supplemented with 50 μg/ml kanamycin, 50 μg/ml spectinomycin, 100 μg/ml ampicillin, and/or 1 mM isopropyl β-D-thiogalactopyranoside (Gold-Bio). Human urine was collected from at least 2 healthy volunteers and filter sterilized through a 0.22-μm filter (Millipore). Biological replicates of growth curves were conducted with different batches of urine collected on different days. Where indicated, urine was supplemented with 10 mM ornithine, citrulline, or arginine (Sigma-Aldrich).

All deletion mutants were generated in the prototypical cystitis isolate UTI89 using the λRed recombinase system (92, 93). The allelic replacement of argI was performed using a previously described, λRed recombinase-based negative-selection system (65). To facilitate *in vivo* competition assays, kanamycin or chloramphenicol resistance markers were inserted into the HK site using the standard λRed recombinase protocol (49, 94). All deletion and allelic replacement mutants were confirmed by Sanger sequencing.

**Growth curves.** Growth curves were acquired on a Tecan infinite M200 Pro (Tecan). Cultures were diluted 1:1,000 into 200 μl of the appropriate medium with arginine, citrulline, or ornithine supplementation where appropriate in CellStar clear polystyrene 96-well flat-bottom plates (Greiner Bio-One) and incubated at 37°C with shaking for 10 to 15 h. Each curve is the average of 4 technical replicates, and the curves are representative of at least 3 independent experiments.

**Hemagglutination.** Hemagglutination assays were performed as described previously (95, 96).

**Mouse models of acute and chronic UTI.** Female, 6- to 7-week-old C3H/HeN HSD mice were acquired from Envigo Laboratories and allowed to acclimate to the Washington University Medical School animal housing facility for 1 week. General infections were performed transurethrally as described previously (11). For single-strain infections, the inoculum contained 2 × 10⁷ to 4 × 10⁷ CFU of a derivative of UTI89 containing a chromosomally integrated kanamycin resistance gene either at the pathogenically neutral lambda attP site (94) or at the site of the desired PSG or arginine synthesis gene. Competitive infections were inoculated with 2 × 10⁷ to 4 × 10⁷ CFU each of a derivative of UTI89 containing a chromosomally integrated spectinomycin resistance gene at the pathogenically neutral HK site (16) and one of the kanamycin resistance (Kmr) gene-containing strains mentioned above, including the wild type to control for pleotropic effects of the kanamycin and spectinomycin resistance genes on fitness during pathogenesis. Allele swap experiments assessing the impact argI variation were performed with UTI89 argI/H11001 KM in the HK site competing against UTI89 argI/MC4100 KM (chloramphenicol resistance gene) in the HK site and with UTI89 argI/H11001 Cmr in the HK site competing against UTI89 argI/MC4100 Cmr in the HK site.

UT tissue burdens were determined as described previously (11). For long-term infections, bacteriuria was monitored as described previously at various time points from 1 to 63 dpi (11). Where appropriate, competitive indices were determined as log₁₀(Spec CFU/Kan CFU)/inoculum (Spec CFU/Kan CFU) or log₁₀(DargI/MC4100 CFU/argI/UTI89 CFU)/inoculum (argI/MC4100 CFU/argI/UTI89 CFU) to control for variations in the initial inoculum.

**Phylogenetic analysis of arginine anabolism.** All RefSeq assemblies listed for *Escherichia coli* (as of 26 April 2016) were downloaded from NCBI GenBank. For a given gene, orthologs were identified using TBLASTN (97); the UTI89 allele of that gene was used as the query against each of the individual assembled genomes. For each genome, only the best single hit that was at least 90% identical over at least 90% of the total length of the UTI89 allele was kept. Unique alleles (based on DNA sequence) were aligned as translated protein sequences; then this protein alignment was imposed on the DNA sequences. Approximate maximum-likelihood trees were created using FastTree 2 (with the –nt and –gtr options).
command line switches) (98). Phylogroups were assigned using an in silico implementation of a triplex PCR (99). All analyses utilized custom Perl scripts. Visualization was done in R (version 3.2.2; https://www.R-project.org) using the ggtree (100), ape (101), and ggplot2 (102) packages.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 2.9 MB.
FIG S2, PDF file, 0.1 MB.
FIG S3, PDF file, 0.4 MB.
FIG S4, PDF file, 0.2 MB.
FIG S5, PDF file, 0.1 MB.
FIG S6, PDF file, 0.2 MB.
FIG S7, PDF file, 0.3 MB.
FIG S8, PDF file, 0.6 MB.
FIG S9, PDF file, 0.3 MB.

TABLE S1, DOCX file, 0.01 MB.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grants AI048689, DK051406, and DK064540 to S.J.H. M.E.H. was partially supported by AI007172-38. S.L.C. was supported by the National Medical Research Council, Ministry of Health, Singapore (grant number NMRC/CIRG/1467/2017).

We thank Matt Conover, Jenn Siverman, Jerry Pinkner, Tom Hannan, and Nicole Riley for their helpful discussion during preparation of the manuscript.


REFERENCES


Positive Selection and Arginine Metabolism in UPEC


